

(19)日本国特許庁(JP)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開平5-282380

(43)公開日 平成5年(1993)10月29日

(51)Int.Cl.<sup>5</sup>

G 0 6 F 15/40

識別記号

5 3 0 Q 7060-5L

庁内整理番号

F I

技術表示箇所

審査請求 未請求 請求項の数1(全 6 頁)

(21)出願番号 特願平4-103977

(22)出願日 平成4年(1992)3月31日

(71)出願人 000004226

日本電信電話株式会社

東京都千代田区内幸町一丁目1番6号

(72)発明者 君山 博之

東京都千代田区内幸町1丁目1番6号 日

本電信電話株式会社内

(72)発明者 清末 悌之

東京都千代田区内幸町1丁目1番6号 日

本電信電話株式会社内

(72)発明者 大庭 有ニ

東京都千代田区内幸町1丁目1番6号 日

本電信電話株式会社内

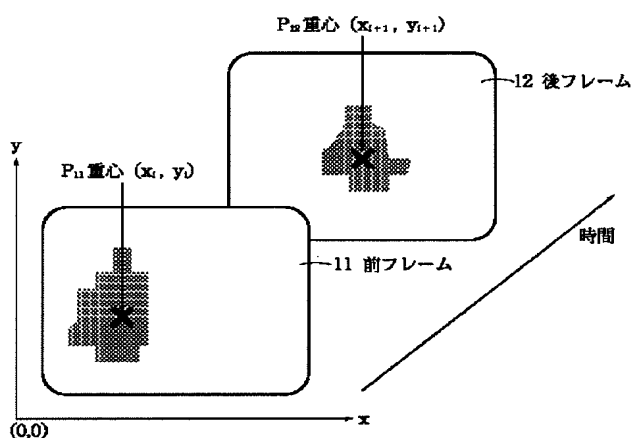
(74)代理人 弁理士 小林 将高

(54)【発明の名称】 動画像検索方法

(57)【要約】

【目的】 フレーム単語で人間に理解可能なインデックスを付与し、そのインデックスを用いて任意の場面を検索できるようにする。

【構成】 動画内容を表示する物理量を複数のカテゴリ、例えば「被写体の大きさ」とか「被写体の色」とかに分類し、このカテゴリを表示するのに複数の単語、例えば「大、中、小」、「赤、黄、緑、……」を入れた単語データベースをあらかじめ作っておき、動画像を処理して得られる物理量をもとに前記単語データベースの中から複数の単語を計算機に自動的に選択させ、これを単語フレーム単位で付与しインデックスとする。検索実行時には、検索したい場面に相当する単語を入力して、一致した動画像のフレームを候補画像とする。



**【特許請求の範囲】**

**【請求項1】** 動画像内容を表示する物理量を複数のカテゴリに分類し、これらのカテゴリを表示する単語が複数個入った単語データベースを用意しておき、動画像を処理して得られる物理量をもとに、前記単語データベースの中から複数の単語を計算機に自動的に選択させ、選択させた複数の単語をフレーム単位で付与し、検索実行時に、ユーザに前記単語データベースから検索したい場面に相当する単語を選択させ、その選択させた単語と同じ単語が付与されている動画像のフレームを候補画像として提示することを特徴とする動画像検索方法。

**【発明の詳細な説明】****【0001】**

**【産業上の利用分野】** 本発明は、動画像データファイルから必要とする場面を検索し、頭出しする際用いる動画像検索方法に関するものである。

**【0002】**

**【従来の技術】** 従来の動画像検索方法は、動画像ファイル単位に手作業で付与されている代表画面を見たり、キーワードを用いて必要な場面の入っているような動画像ファイルを検索し出すファイル単位の検索である。このように、ファイルを発見した後、ファイル中のある場面を検索する場合、ファイルの先頭からそのファイルの中身を見て必要な場面を見つけ出す方法で行われている。また、図8のように、動画像ファイルの代表画面1に輝度分布などの統計量の時間変化図形2と一緒に付与しておき、その統計量を見てファイル中から必要な場面を探す方法がある。

**【0003】**

**【発明が解決しようとする課題】** このような従来の技術では、動画像ファイルという複数フレーム単位でしか検索ができない。前述した代表画面1を用いて細かい単位で検索できるようにするためには、たくさんの代表画面1を一度に見せて、その中からユーザに選択させるという方法をとることになる。しかし、この方法で検索を行う場合、目的の画像を得るために大量の代表画面1をユーザは見なければならないので、大量の時間が必要となる。また、前述したキーワードを利用する場合、フレーム単位でキーワードを付与しなければならないので、付与するのに多くの時間と人手が必要となる。

**【0004】** 本発明の目的は、フレーム単位で人間に理解可能なインデクスを付与する方法とそのインデクスを用いた検索方法を提供することによって任意の場面を検索できるようにすることにある。

**【0005】**

**【課題を解決するための手段】** 本発明にかかる動画像検索方法では、前記課題を解決するために、前処理として動画像内容を表す物理量をあらかじめ複数のカテゴリに分類しておき、そのカテゴリを表示するための基本的な単語を複数定め、単語データベースとして用意してお

く。そして、動画像を単一フレームおよび前後フレーム間で処理することによって生成される物理量をもとに、前述した単語データベースの中から、各カテゴリについて自動的に選択させる。この作業を行った後、フレーム単位にそれらの単語を動画像に付与する。この方法によって、フレーム単位に自動的にインデクスを付与することが可能となる。そして、検索時に、各カテゴリ毎に検索したい場面を表している単語を単語データベースの中から選択させ、選択された単語と同じ単語が付与されているフレームを候補として提示させることによって、フレーム単位に検索するものである。

**【0006】**

**【作用】** 本発明においては、検索時に各カテゴリ毎に検索したい場面を表している単語を直接または間接的に入力するのみでよいので、人間によって行われてきた作業を軽減することが可能となるとともに、動画像ファイル中の任意の場所を検索しようとする場合にも、検索時間が短縮される。

**【0007】**

**【実施例】** 以下に、本発明の動画像検索方法の一実施例を示す。本発明の動画像検索方法は大きく分けて、検索のためのインデクスとなる複数の単語を単語データベースから選択し付与する過程、その単語を用いて検索を行う過程の2つからなる。

**【0008】** 最初に、単語の生成方法の一例を示す。ここでは、単語データベースの一例として、次のものを用意しておく。「被写体の大きさ」、「被写体の色」、「被写体の移動速度」、「被写体の移動方向」をカテゴリとして、それぞれのカテゴリに対して図7の単語を与えておく。すなわち、図7の例では、「被写体の大きさ」のカテゴリには「大、中、小」の3つの単語を用意し、以下、同様に「被写体の色」に対しては「赤、黄、緑、青、紫」を、「被写体の移動速度」に対しては「速い、中速、遅い、静止」を、さらに、「被写体の移動速度」に対しては「右、右上、上、左上、左、左下、下、右下」を用意しておく。

**【0009】** 次に、計算機による単語の割り振り方法について述べる。上記のカテゴリが被写体に関するものであるため、フレーム毎に被写体領域の抽出を行う。被写体領域の抽出方法は、例えば、「動物体領域抽出に関する一検討（1991年秋信学全大、D-200、1991）」等の方法で、先頭フレームについて被写体領域を指定することによって行うことが可能である。これらの方法によって領域抽出を行った後、抽出した領域の重心座標、面積、幅、高さ、色相等の特徴量をそれぞれの領域に対して求める。そして、このようにして求めた物理的な特徴量をもとに単語の割り振りを行う。重心位置の値を用いて被写体の移動速度と移動方向に関する単語を割り振る方法の一例を図1を用いて説明する。はじめに、移動方向の決定方法として、まず、図1における後

フレーム12, 前フレーム11の重心 $P_{12}$ ,  $P_{11}$ を示す座標 $x_{i+1}$ ,  $x_i$ ,  $y_{i+1}$ ,  $y_i$ から下記(1),

(2)式を用いて、 $dx_i$ と $dy_i$ をそれぞれ求め、 $dx_i$ と $dy_i$ の符号によって移動方向の単語を決定する。

【0010】

【数1】

$$dx_i = x_{i+1} - x_i \quad \dots\dots (1)$$

$$dy_i = y_{i+1} - y_i \quad \dots\dots (2)$$

図1の例では、左下を座標原点とすると、 $dx_i > 0$ ,  $dy_i > 0$ となるので図7の「被写体の移動方向」の単語のカテゴリの中から、「右上」という単語を割り振ることができる。移動速度は、(3)式を用いて $dr_i$ を求め、その $dr_i$ の値から「被写体の移動速度」に関する単語、例えば、「中速」を割り振る。

【0011】

【数2】

$$dr_i = dx_i^2 + dy_i^2 \quad \dots\dots (3)$$

以下、同様に、色相データから色に関する単語、例えば、「赤」、面積から大きさに関する単語、例えば、「中」を割り振る。

【0012】次に、これらの単語を用いた検索方法の一例を述べる。図2に以下に述べる検索手順のフローチャートを示す。図2中、(1)～(10)は各ステップを示す。検索手順は、まず、ユーザが検索したい動画のインデックスである単語を指定することから始まる。指定の方法は、直接、単語を指定する方法(1), (2)と、類似場面であるサンプル場면을指示して間接的に単語を指定する方法(1), (9), (10), (2)がある。単語を指定した後、ステップ(2), (3), (4)の検索を開始する。データベース内の動画に付与されている単語と、ユーザが指定した単語が一致すれば(5), (6)、その場면을候補の場面としてユーザに指示し(7)、その中に検索者の意図する場面が含まれていれば(8)、検索を終了する。

【0013】図3に単語を指定するためのユーザインタフェースの一例を示す。図3において、21は単語列指定ウインドウ、22は検索実行ボタン、23はサンプル場面表示ウインドウ、24は検索結果表示ウインドウ、25は検索結果確認ウインドウである。図3のインタフェースを用いて直接単語を指定する方法を以下に述べる。

【0014】図4は、図3における単語列指定ウインドウ21の拡大図である。図4において、211はポインティングカーソル、212は単語指定のためのボタン表示を示す。直接、単語を指定する場合、図4のように、単語列指定ウインドウ21中の検索したい場面を表現している単語のアイコンの上にマウス等を用いてカーソル

を移動し、アイコンをクリックすることによって単語の指定を行う。また、類似場面を用いた単語指定方法は、図3のサンプル場面表示ウインドウ23と前述した単語列指定ウインドウ21の2つのユーザインタフェースを用いて行う。

【0015】図5は、図3におけるサンプル場面表示ウインドウ23の拡大図であり、231はコントロールパネル、232はファイル名指定ウインドウ、233はサンプル場面、234はスクロールバー、235は制御ボタンを示す。

【0016】まず、図5のサンプル場面表示ウインドウ23中のファイル名指定ウインドウ232でファイル名を指定し、コントロールパネル231中の再生ボタンや早送りボタン、スクロールバー234でサンプル場面233を表示させると、その場面对応する単語が図4の単語列指定ウインドウ21に表示される。この単語列指定ウインドウ21に表示された内容が検索者の意図する場合と同様の方法で修正することが可能である。

【0017】以上述べたように、単語を指定した後、図3の中の検索実行ボタン22上にマウスなどを用いてカーソルを移動させ、その上でボタンをクリックすることによって検索を開始する。検索方法は、単語列指定ウインドウ21に表示されている単語と、画像データベース中の画像に前述した方法であらかじめ付与しておいた単語とを比較し、単語が一致した場면을図3のように検索結果表示ウインドウ24に表示する。図3の検索結果表示ウインドウ24に表示された場면을、さらに詳しく見たい場合、検索結果表示ウインドウ24中の見たい場面上にカーソルをマウスなどで移動し、マウスのボタンをクリックすることによってクリックした場面が、図3の検索結果確認ウインドウ25に表示される。検索結果確認ウインドウ25の拡大図を図6に示す。図中の再生ボタンや早送りボタン等によって結果の確認を行うことができる。もし、検索結果表示ウインドウ24に表示された場面が意図した場面と異なる場合、図4のように単語列指定ウインドウ21で再度単語を指定し、検索を実行し直すことによって目的の場面を検索することができる。

【0018】

【発明の効果】以上詳細に説明したように、本発明は、動画内容を表示する物理量を複数のカテゴリに分類し、これらのカテゴリを表現する単語が複数個入った単語データベースを用意しておき、動画を処理して得られる物理量をもとに、単語データベースの中から複数の単語を計算機に自動的に選択させ、選択させた複数の単語をフレーム単位で付与し、検索実行時に、ユーザに単語データベースから検索したい場面に相当する単語を選択させ、その選択させた単語と同じ単語が付与されている動画のフレームを候補画像として提示するように

したので、従来まで手作業で行ってきたインデクシング作業の大部分を計算機に行わせることができるとともに、任意の単位（例えばフレーム単位）でインデクスを付与できるため、動画ファイル中の任意の場面を検索することが可能となる。

【図面の簡単な説明】

【図1】本発明における単語生成方法の説明図である。

【図2】本発明における画像検索のフローチャートの例を示す図である。

【図3】本発明における画像検索のためのインタフェース画面の例を示す図である。

【図4】本発明における単語列指定ウインドウの例を示す図である。

【図5】本発明におけるサンプル場面表示ウインドウの例を示す図である。

【図6】本発明における検索結果確認ウインドウの例を示す図である。

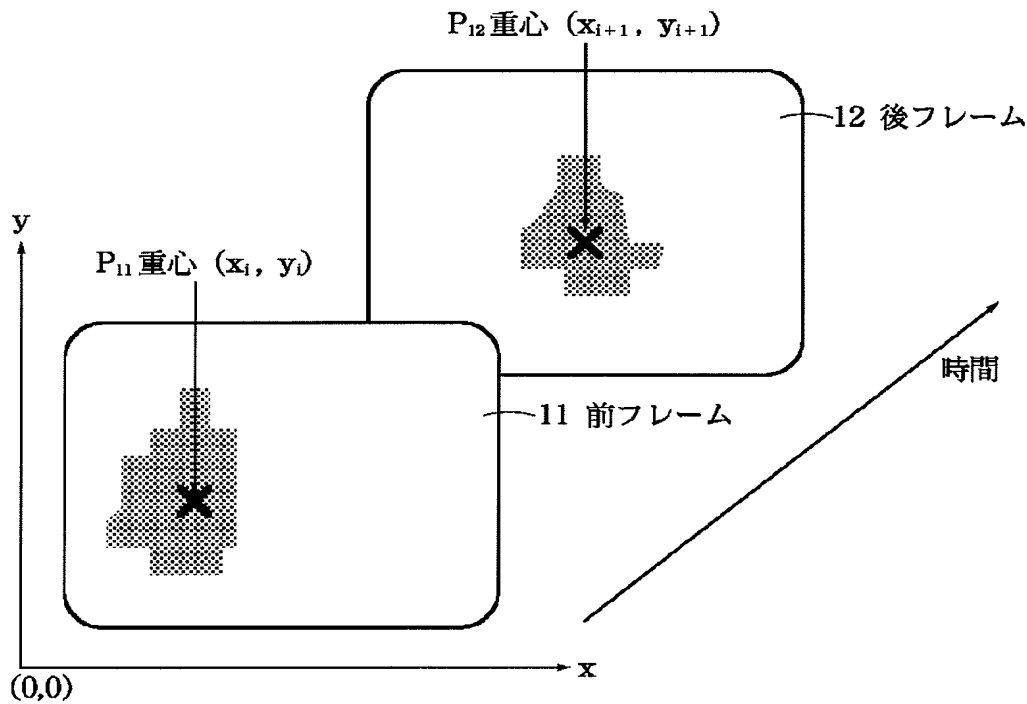
【図7】本発明における被写体の内容表現する物理量のカテゴリと、これらのカテゴリを表現する単語とからなるデータベースの一例を示す図である。

【図8】従来の統計量を用いた動画表示の例を示す図である。

【符号の説明】

- 11 前フレーム
- 12 後フレーム
- 21 単語列指定ウインドウ
- 22 検索実行ボタン
- 23 サンプル場面表示ウインドウ
- 24 検索結果表示ウインドウ
- 25 検索結果確認ウインドウ

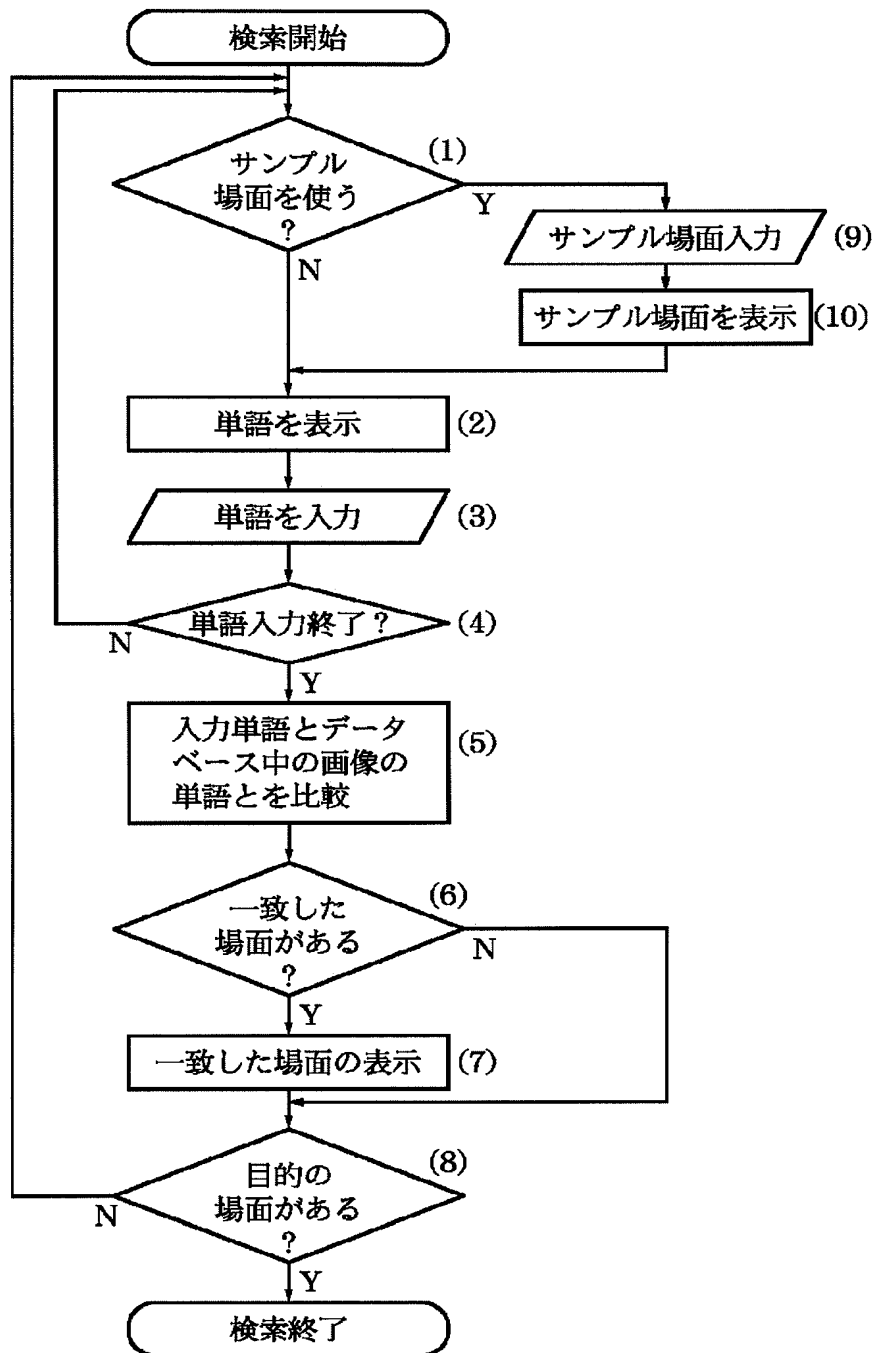
【図1】



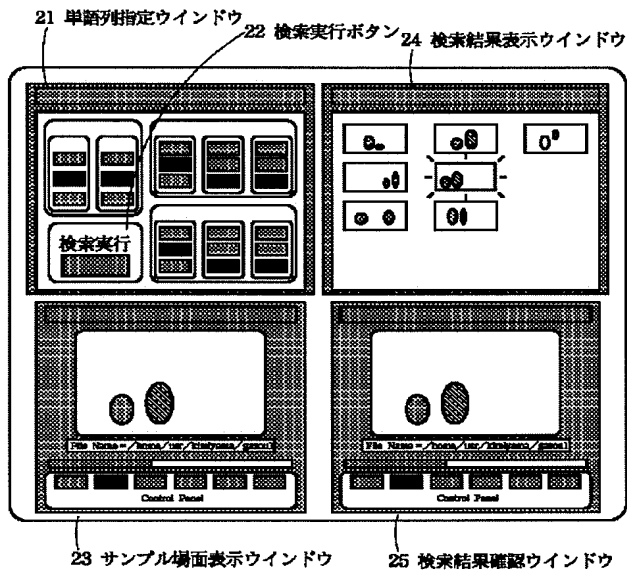
【図7】

単語のカテゴリ	単語
被写体の大きさ	大、中、小
被写体の色	赤、黄、緑、青、紫
被写体の移動速度	速い、中速、遅い、静止
被写体の移動方向	右、右上、上、左上、左、左下、下、右下

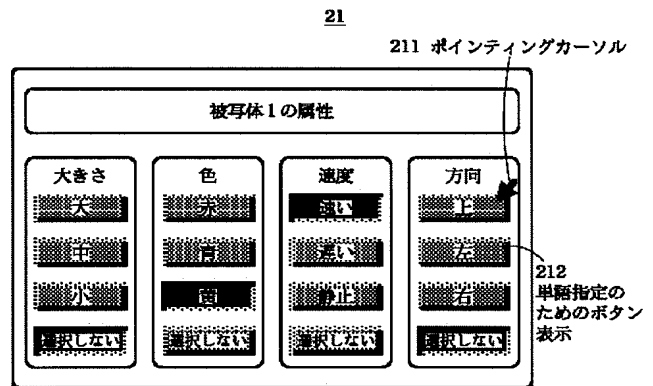
【図2】



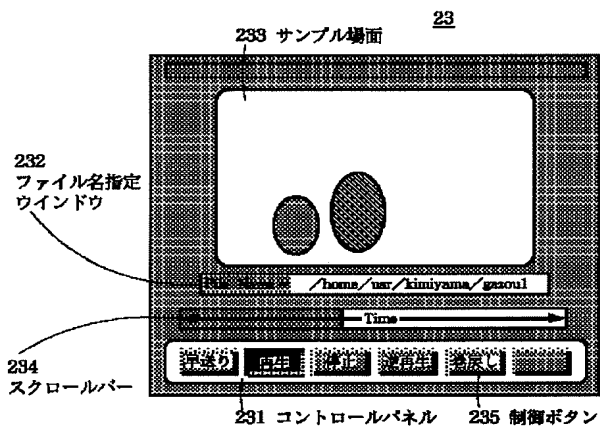
【図 3】



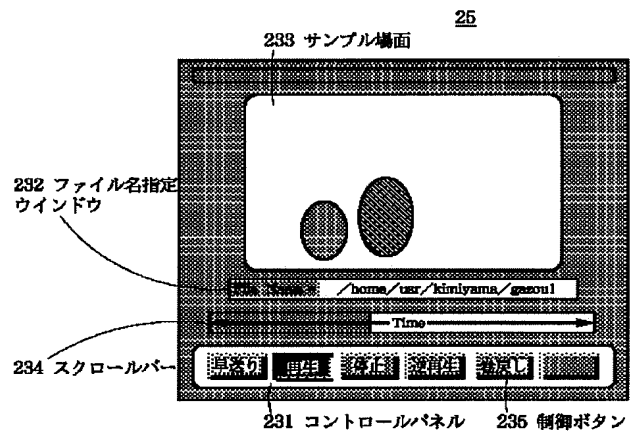
【図 4】



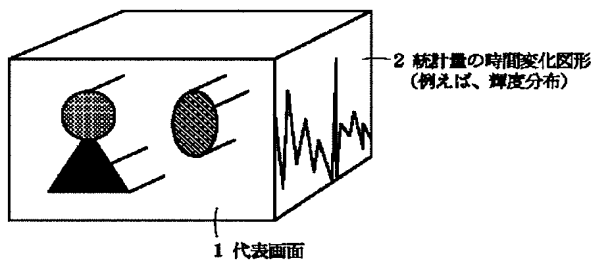
【図 5】



【図 6】



【図 8】



# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 05-282380

(43)Date of publication of application : 29.10.1993

---

(51)Int.Cl.

G06F 15/40

---

(21)Application number : 04-103977 (71)Applicant : NIPPON TELEGR & TELEPH  
CORP <NTT>

(22)Date of filing : 31.03.1992 (72)Inventor : KIMIYAMA HIROYUKI  
KIYOSUE TOMOYUKI  
OBA YUJI

---

## (54) ANIMATION IMAGE RETRIEVING METHOD

### (57)Abstract:

PURPOSE: To retrieve an optional scene by means of an index which can be understood by humans in a frame word.

CONSTITUTION: The physical value showing the animation contents is sorted into plural categories e.g. the size and the color of an object etc. A word data base including plural words like 'large' 'medium' 'small' 'red' 'yellow' 'green' etc. is prepared for expression of the preceding categories. Then plural words are automatically selected by a computer out of the word data base based on the physical value obtained through the processing of an animation. These selected words are applied as the indexes for each word frame. In a retrieving execution state the word corresponding to the desired scene to be retrieved is inputted. Then the frame of the coincident animation image is defined as a candidate image.

---

## CLAIMS

---

### [Claim(s)]

[Claim 1] Classify physical quantity expressing the contents of video into two or more categories and a word data base containing two or more words expressing these categories is prepared. A computer is made to choose two or more words automatically out of said word data base based on physical quantity produced by processing video. A video search method showing a frame of video which makes a word which is equivalent to a scene to search from said word data base to a user at the time of retrieval execution choose and in which two or more words made to choose are given per frame and the same word as the word made to choose is given as a candidate image.

---

## DETAILED DESCRIPTION

---

[Detailed Description of the Invention]

[0001]

[Industrial Application] This invention relates to the video search method used when searching and pulling out the head of the scene needed from a dynamic image data file.

[0002]

[Description of the Prior Art] The conventional video search method is search of the file basis which discovers the video file in which the required scene is likely to be contained using a keyword in seeing the representation screen manually given to the video file basis \*\*\*. Thus when searching a certain scene in a file after discovering a file it is carried out by the method of seeing the contents of the file from the head of a file and finding out a required scene. Like drawing 8 the temporal change figure 2 of statistics values such as luminance distribution is given to the representation screen 1 of a video file together and there is a method of seeing the statistics value and looking for a required scene out of a file.

[0003]

[Problem(s) to be Solved by the Invention] In such a Prior art search is possible only in the multiple frame unit of a video file. In order to enable it to refer to a fine unit using the representation screen 1 mentioned above many representation screens 1 will be shown at once and the method of making a user choose from the inside will be taken. However since the user has to look at a lot of representation screens 1 in order to acquire the target picture when referring to this method a lot of time is needed. Since a keyword must be given per frame when using the keyword mentioned above many time and helps are needed for giving.

[0004] The purpose of this invention is enabling it to search arbitrary scenes by providing the method of giving the index which can be understood to human being per frame and the search method using the index.

[0005]

[Means for Solving the Problem] Physical quantity which expresses the contents of video with a video search method concerning this invention as pretreatment in order to solve said SUBJECT is beforehand classified into two or more categories two or more fundamental words for expressing the category are defined and it prepares as a word data base. And it is made to choose automatically about each category from word data bases mentioned above based on physical quantity generated by processing video by a single frame and order inter-frame. After doing this work those words are given to video per frame. This method enables it to give an index automatically per frame. And it searches per frame by making a frame to which the same word as a word which made choose a word showing a scene to search for every category from word data bases and was chosen is given show as a candidate at the time of search.

[0006]



[Function]The word showing a scene to search with this invention for every category at the time of search may only be inputted directly or indirectly. Thereforesearch time is shortened also when it is going to search the arbitrary places in a video filewhile becoming possible to reduce the work done by human being.

[0007]

[Example]Belowone example of the video search method of this invention is shown. The video search method of this invention is roughly dividedand consists of twothe process in which choose two or more words used as the index for search from a word data baseand they are givenand the process in which it searches using the word.

[0008]Firstan example of the generation method of a word is shown. Herethe following are prepared as an example of a word data base. The word of drawing 7 is given to each category by making "the size of a photographic subject"the color of a photographic subjectthe movement speed of a photographic subjectand the "move direction of a photographic subject" into a category. Namelyin the example of drawing 7three words of small" are prepared for the category of size" of "photographic subject into "sizeHereafterto "the color of a copy body"the right, the upper right, a top, the upper left, the left, the lower left, the bottom, and the lower rightare further prepared [ "red yellowgreenblueand purple" ] for "a quick medium speed and late stillness" to "movement speed of a photographic subject" to "movement speed of a photographic subject" in a similar manner.

[0009]Nexthow to assign the word by a computer is described. Since the above-mentioned category is related with a photographic subjecta photographic subject field is extracted for every frame. The extraction methods of a photographic subject field are methodssuch as "1 examination (the 1991 Akinobu \*\*\*\* sizeD-200-1991) about moving object field extraction"for exampleand can be carried out by specifying a photographic subject field about a head frame. After performing field extraction by these methodscharacteristic quantitysuch as barycentric coordinates of the extracted fieldareawidthheightand hueis calculated from each field. And a word is assigned based on the physical characteristic quantity which was carried out in this way and calculated. An example of the method of assigning the word about the movement speed and the move direction of a photographic subject using the value of a centroid position is explained using drawing 1. the start -- movement -- a direction -- a deciding method -- \*\*\*\*\* -- first -- drawing 1 -- it can set -- a rear frame -- 12 -- a previous frame -- 11 -- the center of gravity --  $P_{12}$  --  $P_{11}$  -- being shown -- coordinates --  $x_{i-1}$  --  $x_i$  --  $y_{i-1}$  --  $y_i$  -- from -- the following -- (one --). (2) Using a formulaask for  $dx_i$  and  $dy_i$  respectively and the numerals of  $dx_i$  and  $dy_i$  determine the word of the move direction.

[0010]

[Equation 1]

In the example of drawing 1 if the lower left is made into an origin of coordinates since it will be set to  $dx_i > 0$  and  $dy_i > 0$  the word the "upper right" can be assigned out of the category of the word of the "move direction of a photographic subject" of drawing 7. Movement speed asks for  $dr_i$  using (3) types and assigns the word about "movement speed of a photographic subject" for example medium speed from the value of the  $dr_i$ .

[0011]

[Equation 2]

Hereafter the word about a color for example red and the word about area to a size for example inside are similarly assigned from hue data.

[0012] Next an example of the search method using these words is described. The flow chart of the retrieving procedure described below is shown in drawing 2. (1) - (10) shows each step among drawing 2. A retrieving procedure begins from specifying first the word which is an index of the video which a user wants to search. The appointed method has directly the method (1) of specifying a word (2) and the method (1) of pointing to the sample scene which is a similar scene and specifying a word indirectly (9) (10) and (2). After specifying a word search of a step (2) (3) and (4) is started. (8) and search will be ended if the scene which will point to (5) (6) and its scene to a user as a candidate's scene and a retrieving person will mean (7) and in it if the word given to the video in a database and the word specified by a user are in agreement is included.

[0013] An example of the user interface for specifying a word is shown in drawing 3. As for a retrieval execution button and 23 in drawing 3 21 is [ a search-results display window and 25 ] search-results check windows a sample scene display window and 24 the word sequence appointed window and 22. How to specify a word directly using the interface of drawing 3 is described below.

[0014] Drawing 4 is an enlarged drawing of the word sequence appointed window 21 in drawing 3. In drawing 4 211 shows pointing cursor and 212 shows the button display for word specification. When specifying a word directly onto the icon of the word which is expressing the scene in the word sequence appointed window 21 to search like drawing 4 a mouse etc. are used cursor is moved and a word is specified by clicking on an icon. The word specification method using a similar scene is performed using two user interfaces the sample scene display window 23 of drawing 3 and the word sequence appointed window 21 mentioned above.

[0015] Drawing 5 is an enlarged drawing of the sample scene display window 23 in drawing 3.

The file name appointed window and 233 show a sample scene 234 shows a scroll bar and 232231 shows a control button 235.

[0016] First a file name is specified in the file name appointed window 232 in the sample scene display window 23 of drawing 5 If the sample scene 233 is displayed

with the reproduction button and fast forward button in the control panel 231 and the scroll bar 234 the word corresponding to the scene will be displayed on the word sequence appointed window 21 of drawing 4. When the contents displayed on this word sequence appointed window 21 differ from the contents which a retrieving person means it is possible to correct by the method same like drawing 4 as the case where a word is specified directly.

[0017] As stated above after specifying a word on the retrieval execution button 22 in drawing 3 a mouse etc. are used cursor is moved and search is started by clicking a button on it. A search method compares the word currently displayed on the word sequence appointed window 21 with the word beforehand given by the method mentioned above in the picture in an image database and displays the scene whose word corresponded on the search-results display window 24 like drawing 3. When you would like to see in more detail the scene displayed on the search-results display window 24 of drawing 3 cursor is moved with a mouse etc. onto the scene in the search-results display window 24 to see and the scene clicked by clicking the button of a mouse is displayed on the search-results check window 25 of drawing 3. The enlarged drawing of the search-results check window 25 is shown in drawing 6. A reproduction button a fast forward button etc. in a figure can perform the check of a result. When it differs from the scene which the scene displayed on the search-results display window 24 meant the target scene can be searched by specifying a word again in the word sequence appointed window 21 like drawing 4 and reperforming a search.

[0018]

[Effect of the Invention] As explained to details above the physical quantity expressing the contents of video is classified into two or more categories according to this invention. The word data base containing two or more words expressing these categories is prepared. A computer is made to choose two or more words automatically out of a word data base based on the physical quantity produced by processing video. Give two or more words made to choose per frame the word which is equivalent to a scene to search from a word data base to a user at the time of retrieval execution is made to choose and the frame of the video in which the same word as the word made to choose is given was shown as a candidate image.

Therefore since an index can be given in arbitrary units (for example frame unit) while being able to make the great portion of indexing work manually done to the former do on a computer it becomes possible to search the arbitrary scenes in a video file.

---

## DESCRIPTION OF DRAWINGS

---

[Brief Description of the Drawings]

[Drawing 1] It is an explanatory view of the word generation method in this

invention.

[Drawing 2] It is a figure showing the example of the flow chart of the image retrieval in this invention.

[Drawing 3] It is a figure showing the example of the interface screen for the image retrieval in this invention.

[Drawing 4] It is a figure showing the example of the word sequence appointed window in this invention.

[Drawing 5] It is a figure showing the example of the sample scene display window in this invention.

[Drawing 6] It is a figure showing the example of the search-results check window in this invention.

[Drawing 7] It is a figure showing an example of the database which consists of a category of the physical quantity expressing the contents of the photographic subject in this invention and a word expressing these categories.

[Drawing 8] It is a figure showing the example of the cine mode display using the conventional statistics value.

[Description of Notations]

11 Previous frame

12 Rear frame

21 Word sequence appointed window

22 Retrieval execution button

23 Sample scene display window

24 Search-results display window

25 Search-results check window

---